CLAIMS

1. A method of providing a representative, non-redundant overview of the peptide content of a sample type by analyzing a plurality of samples using its peptide topology, wherein the method comprises the steps of:

a) providing a respective mass spectrum for each sample of said plurality of samples, wherein signal intensity peaks correspond to potential peptides,

b) computing the measures of correlation between the signal intensities of said potential peptides,

c) grouping potential peptides together, which exhibit a degree of correlation among each other above a certain threshold, thereby providing a plurality of correlation associated networks of potential peptides, and

d) assigning at least one representative potential peptide out of at least one correlation associated network as a representative peptide to said correlation associated network of said sample type.

2. A method for predicting the sequence of peptides using the peptide topology of a plurality of samples containing a peptide having a known precursor, wherein the method comprises the steps of:

a) providing a respective mass spectrum for each sample of said plurality of samples, wherein signal intensity peaks correspond to potential peptides,

b) identifying said peptide having a known precursor using the mass of said peptide, wherein the sequence of the known precursor is known,

c) computing the measures of correlation between the signal intensity of said peptide having a known precursor and the signal intensities of the other potential peptides,

d) selecting potential peptides, which exhibit a degree of correlation with said peptide having a known precursor above a certain threshold, and

e) predicting the sequence of the potential peptides by matching masses of putative fragments of the sequence of the known precursor with the masses of the potential peptides correlating with said peptide having a known precursor.

3. A method for predicting the sequence of peptides using the peptide topology of a plurality of samples containing a peptide with a known sequence, wherein the method comprises the steps of:

a) providing a respective mass spectrum for each sample of said plurality of samples, wherein signal intensity peaks correspond to potential peptides,

b) identifying a peptide with a known sequence using its mass,

c) computing the measures of correlation between the signal intensity of said known peptide and the signal intensities of the potential peptides,

d) selecting potential peptides, which exhibit a degree of correlation with the known peptide above a certain threshold,

e) computing the mass differences between each of the potential peptides and the known peptide, and

f) predicting the sequence and/or the biologically, chemically or physically modified sequence of the potential peptides by using data about mass differences caused by biological, chemical or physical processes matching the mass differences determined in step e).

4. A method for identifying peptides suitable to be used as marker panels using the peptide topology of a plurality of samples taken from at least two different experimental groups representing a status A and a status B, wherein the method comprises the steps of:

a) providing a respective mass spectrum for each sample of said plurality of samples, wherein signal intensity peaks correspond to potential peptides,

b) computing the measures of correlation between the signal intensities of said potential peptides for each plurality of samples within each experimental group separately, and

c) selecting pairs of potential peptides, which exhibit a difference in the degree of correlation between the different experimental groups above a certain threshold, thereby providing peptides which are suitable to be used as marker panels for diagnostic purposes to distinguish between status A and status B.

5. A method for identifying peptides suitable to be used as marker panels using the peptide topology of a plurality of samples taken from at least two different experimental groups representing a status A and a status B, wherein the method comprises the steps of:

a) providing a respective mass spectrum for each sample of said plurality of samples, wherein signal intensity peaks correspond to potential peptides,

b) selecting potential peptides correlating with a parameter being representative of status A or status B,

c) computing the measures of correlation between the signal intensities of said selected potential peptides for each plurality of samples, and

d) selecting pairs of potential peptides which exhibit no correlation of their respective signal intensities above a certain threshold, thereby providing potential peptides which are suitable to be used as complementing peptides in a marker panel for diagnostic purposes to distinguish between status A and status B.

6. A method for identifying peptides suitable as a surrogate for a known peptide using the peptide topology of a plurality of samples, wherein the method comprises the steps of:

a) providing a respective mass spectrum for each sample of said plurality of samples, wherein signal intensity peaks correspond to potential peptides,

b) computing the measures of correlation between the signal intensity of said known peptide and the signal intensities of potential peptides, and

c) selecting potential peptides, which exhibit a degree of correlation with said known peptide above a certain threshold, thereby providing potential peptides suitable as a surrogate for said known peptide.

7. The method according to any one of claims 1 to 3 or 6, where a plurality of minimal spanning tree diameters is computed as a measure of correlation using the signal intensity of said potential peptides in said samples, wherein the selection of potential peptides is done by using minimal spanning tree diameter threshold, wherein the minimal spanning tree diameter for an association of two potential peptides has to be above an adjustable threshold of at least 0.425 times the number of samples.

8. The method according to claims 4 or 5, where a plurality of minimal spanning tree diameters is computed as a measure of correlation using the signal intensity of said potential peptides in said samples, wherein the selection of pairs of potential peptides is done by using minimal spanning tree diameter threshold, wherein the difference between the minimal spanning tree diameter found in the said different experimental groups is above an adjustable threshold of at least 0.1 times the number of samples.

9. The method according to any one of the preceding claims, wherein the method comprises the additional step of at least one fractionating step of said samples prior to providing the mass spectra of said samples and wherein at least one fraction of said samples is used for providing said mass spectra.

10. The method according to any one of the preceding claims using at least one measure of correlation selected from the group consisting of "Pearson Product-Moment Correlation Coefficient", "Spearman's rank order Correlation Coefficient", "Kendall's Tau", "Kendall's Coefficient of Concordance", "Goodman and Kruskal's Gamma" and "Minimal Spanning Tree diameters".

11. The method according to any one of the preceding claims using at least one method for calibrating the mass spectrometric data selected from the group consisting of "Simple Offset Correction", "2-Point Baseline Correction", "Multi-Point Baseline Correction", "Interactive Polynomial Baseline Correction", "Function Fit Baseline Correction", and "GIFTS (Auto Leveling Method) Baseline Correction".

12. The method according to any one of the preceding claims using at least one method for identifying outlier samples selected from the group consisting of "Principal Component Analysis", "multivariate calibration partial least-squares", and "Replicator Neural Networks".

13. The method according to any one of the preceding claims, wherein the calculation of the measures of correlation is repeated at least once using the peptide coordinates resulting from the previous round of calculations of measures of correlation, thereby providing the measures of correlation of 2^{nd} or higher order neighborhood.

14. The method according to any one of the preceding claims using additional coordinates besides the mass selected from the group consisting of fraction number, elution time, retention time, protein chip coordinates, peptide concentration, enzyme activities, structural properties, chemical properties and biological properties.

15. The method according to any one of the preceding claims, wherein MALDI mass spectrometry or ESI mass spectrometry is used to generate the mass spectra.

16. The method according to any one of the preceding claims, wherein the samples or groups of samples are homogeneous.

17. The method according to any one of the preceding claims, wherein the computation of measures of correlation is done in advance prior to the analysis to accelerate the speed of the analysis using pre-determined values the measures of correlation.

18. The method according to any one of the preceding claims, wherein the necessary sequence information is provided by manual input or automatically queried from a database.

19. The method according to any one of the preceding claims, wherein the corresponding results are automatically combined with data from other sources chosen from the group consisting of sequence databases, patent databases, literature databases, medical databases, 3D structure databases, databases containing information about enzyme recognition sites, posttranslational modifications, genetic polymorphisms, clinical trials.

20. The method according to any one of the preceding claims, wherein at least one step of data processing or data supply is done using a remote computer system and wherein the user is connected via an internet, intranet or other network to the remote computer system.

21. A digital computer system programmed to perform a method according to any one of the preceding claims.

22. A computer readable medium storing a computer program implementing a method according to any one of claims 1 to 20.

23. Use of a method according to any one of the previous claims, wherein at least part of the data-analysis is done via a remote computer system located in a different country.

24. Use of a method according to any one of claims 2, 3 and 6 to 23 for determining alterations in the amino acid sequence length and/or for determining chemical or posttranslational modifications of peptides of known identity which peptides were added to the sample.